Current Biology 18, 1760–1765, November 25, 2008 ©2008 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2008.09.059

Report

Evidence for Insertional RNA Editing in Humans

Alexandre Zougman,[1,3] Piotr Ziółkowski,[2] Matthias Mann,[1,*] and Jacek R. Wiśniewski[1,*]
[1]Department of Proteomics and Signal Transduction
Max Planck Institute for Biochemistry
Am Klopferspitz 18
D-82152 Martinsried
Germany
[2]Department of Pathology
Wroclaw Medical University
ul. Marcinkowskiego 1
PL-50-368 Wroclaw
Poland
[3]Center for Integrated Protein Science
D-81377 Munich
Germany

Summary

Large-scale analysis directly at the protein level holds the promise of uncovering features not apparent or present at the gene level [1–3]. Although mass spectrometry (MS)-based proteomics can now identify and quantify thousands of cellular proteins in large-scale proteomics experiments, much of the peptide information contained in these experiments remains unassigned [4]. Here, we use such information to discover a previously unreported mechanism creating altered protein forms. Linker histones H1 and high-mobility group (HMG) proteins are abundant nuclear proteins that regulate gene expression through modulation of chromatin structure [5–8]. In the high-resolution MS analysis of histone H1 and HMG protein fractions isolated from human cells, we discovered peptides that mapped upstream of the known translation start sites of these genes. No alternative upstream start site exists in the genome, but analysis of Expressed Sequence Tag (EST) databases revealed that these N-terminally extended (ET) proteins are due to in-frame translation of the 5′ untranslated region (5′UTR) sequences of the transcripts. The new translation start sites are created by a single uridine insertion between AG, reflecting a previously unreported RNA-editing mechanism. To our knowledge, this is the first report of RNA-insertion editing in humans and may be an example of the type of discoveries possible with modern proteomics methods.

Results and Discussion

Identification of N-Terminally Extended (ET) Proteins

In-depth proteomic analysis of nuclear extracts from various human cell lines revealed a multitude of posttranslational modifications in linker histone H1 and high-mobility group (HMG) proteins [9]. Apart from this, we found a number of fragmentation spectra of excellent quality that could not be matched to any protein sequence or open reading frame (ORF) in the human International Protein Index (IPI) database. We therefore extracted partial de novo sequences from these fragmentation spectra. When searching the NCBI dbEST database [10] with these sequences, we found that they matched directly upstream of the start codon and within the 5′UTR of *h1.0.* Similarly, we also found peptides pertaining to the 5′UTR of the *hmgn1* gene. This was surprising because in both cases, the 5′UTRs do not contain any alternative start codons.

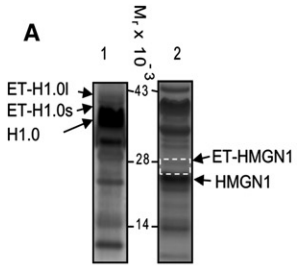
To characterize the N-terminally extended (ET) sequences of H1.0 and HMGN1, we partially purified these proteins from MCF7 cells by reversed-phase chromatography, resolved fractions by SDS PAGE (Figure 1A), digested them with trypsin, and analyzed peptide mixtures by online liquid-chromatography mass spectrometry (LCMS). We identified a total of eight unique peptides (including one phosphopeptide) of ET-H1.0 and two unique peptides of ET-HMGN1 using high-resolution, high-mass-accuracy MS. Several of these peptides were further verified by high-resolution fragmentation analysis with low ppm accuracy (see Appendices S1 and S2, available online). Fragmentation spectra of the identified peptides containing the "normal" initiation methionine, as well as the peptides most proximal to the N termini of ET-H1.0 and ET-HMGN1, are shown in Figure 1. We identified two forms of ET-H1.0, visible as two bands in SDS PAGE, and a single form of ET-HMGN1 (Figure 1F). On the basis of the mapped peptides, the long and short ET extensions of H1.0 comprise at least 86 and 63 amino acid residues, respectively, whereas the ET extension of HMGN1 is at least 23 amino acid residues in length. Due to their amino acid composition, the high number of the basic residues, and the low number of the hydrophobic residues, histones have reduced mobility in SDS PAGE. With respect to the molecular weight (MW) markers, H1 histones appear at a MW of 30,000–35,000 Da. However, their molecular masses are in the range of 20,000–22,000 Da. This is the reason that the ET-H1.0 forms appear to have MWs close to 40,000 Da.

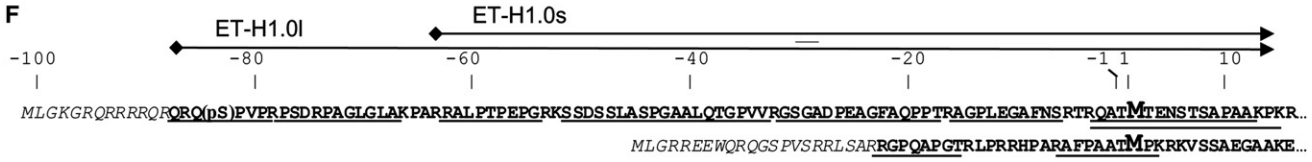
New Translation Start Sites Are Created by a Single Uridine Insertion

We then searched the NCBI dbEST database for EST fragments containing the 5′UTR of *h1.0* and found 11 sequences, derived from different tissue sources, with a uridine (U) insertion (the NCBI dbEST database release of August 06, 2008 contains 301 ESTs carrying the *h1.0* 5′UTR, 11 of them with U insertion). In each case, this insertion happened at the same position—297 bp upstream of the known start codon—between the A and G bases at the AGCT location (Figure 2B, Appendices S3, S4, and S5). For *hmgn1*, we likewise found an EST sequence in which a new start codon is potentially created by U insertion, in this case 135 nt upstream of the known start codon (Appendices S3 and S5). With these start sites, the identified peptides cover 85% and 50% of the predicted sequences of ET-H1.0 and ET-HMGN1, respectively (Figure 2A).

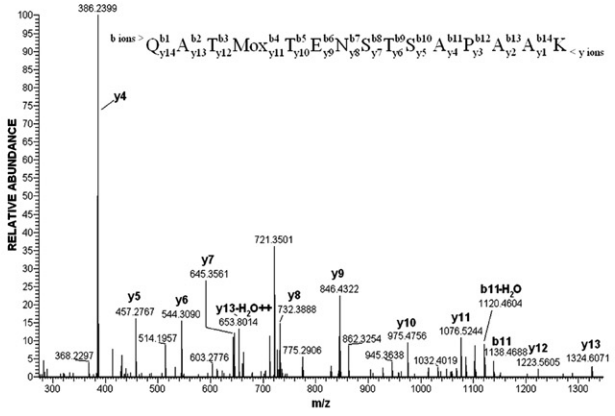
h1.0 occurs as a single gene in the genome database. To exclude the existence of a genomic form of ET-*h1.0*, we performed PCR analysis of genomic DNA for the sequence suggested by the EST data. The "normal" *h1.0* gene has an *AluI* restriction site (5′-AG/CT-3′) in its 5′UTR sequence at the position where the U insertion was found in some EST sequences

*Correspondence: mmann@biochem.mpg.de (M.M.), jwisniew@biochem.mpg.de (J.R.W.)

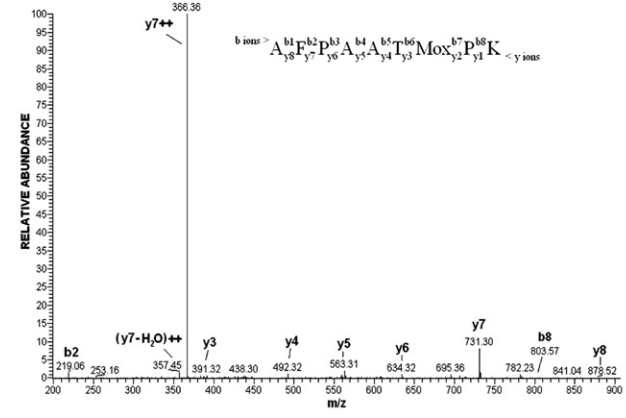




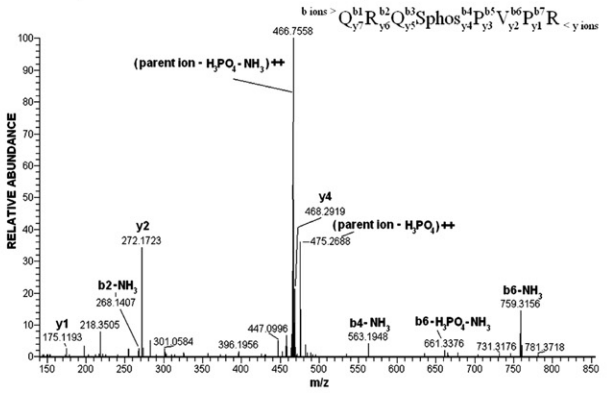
B Histone ET-H1.0 (residues -3−12)



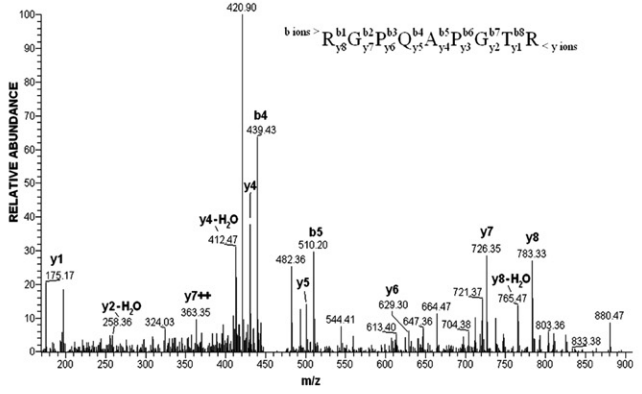
C High Mobility Group Protein ET-HMGN1 (residues -6−3)



D Histone ET-H1.0 (residues -86−-79)



E High Mobility Group Protein ET-HMGN1 (residues -23−18)



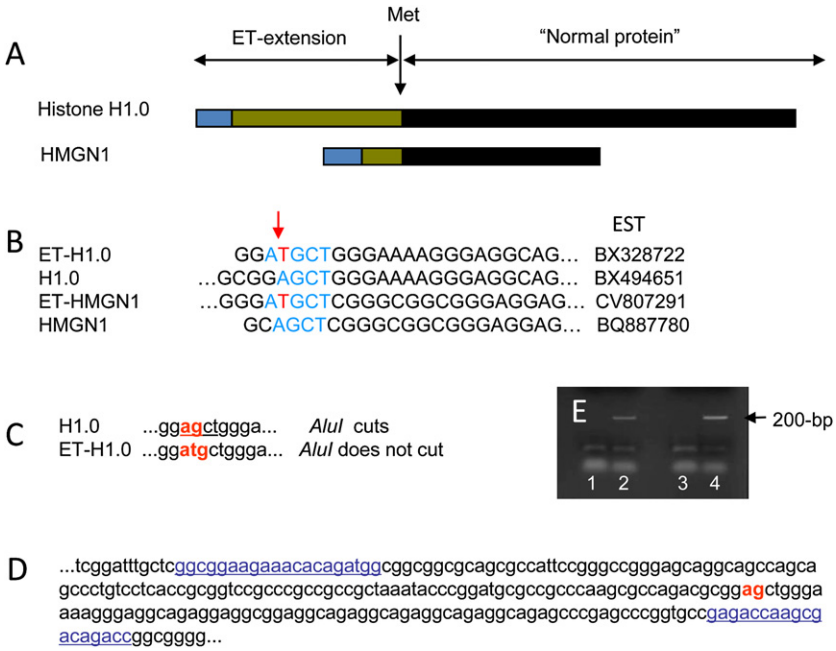


Figure 2. Primary Structure and Putative Origin of ET Proteins

(A) Schematic view of the proteins with extension of their termini (ET) originating from translation of the 5′UTRs. Black, canonical translation products; green, sequenced by high-resolution mass spectrometry; blue, predicted primary structure of the ET-H1.0 and ET-HMGN1 proteins.

(B) AGCT site with the T insertion (arrow) in EST sequences coding for ET proteins.

(C) The *AluI* site of the "normal" *h1.0* sequence is impaired by the creation of the new start codon.

(D) The 200 bp sequence containing the base insertion site (red bold) was amplified with genomic DNA isolated from MCF7 cells. The sequences of the used primers are underlined.

(E) The 200 bp PCR product was obtained only from *AluI*-undigested templates. Lane1, the DNA template was digested with *AluI* prior to the PCR reaction; Lane 2, the PCR reaction with intact template; Lanes 3 and 4, the 200-bp product from lane 2 was extracted from the gel and incubated in the presence or absence of *AluI*, respectively.

(Figure 2C). The insertion creates a new start codon and impairs the *AluI*-restriction site. Thus, amplification of the 200 bp sequence (Figure 2D) from *AluI*-treated DNA would allow identification of the modified fragment, and the PCR product obtained from the undigested template would be susceptible to *AluI* digestion. The PCR reactions revealed that neither was the 200 bp fragment amplified from the *AluI*-digested template nor was the 200 bp PCR product from the undigested template cleavable by the endonuclease (Figure 2E). These results disprove the presence of *h1.0* gene sequence coding for N-terminally extended H1.0, confirming that the new start codon is created during or after transcription of *h1.0* by an RNA editing process. It is possible that the *hmgn1* mRNA is processed by the same mechanism as that of *h1.0*. While RNA insertion has not been described in metazoans, it is not completely unprecedented in biology and may resemble the RNA editing mechanism described for trypanosomatid protozoans in which the uridine insertion/deletion-based RNA editing of the mitochondrial mRNA creates new initiation/termination codons [11]. The resulting N-terminal protein sequences can carry unique properties such as, for example, DNA binding [12]. The human 5′UTRs of H1.0 and HMGN1 have no obvious similarities with uridine insertion regions in the trypanosomatid mitochondrial genes. These regions also are not conserved between species.

ET Proteins Occur Abundantly in Human Cells

To characterize ET-H1.0 and ET-HMGN1 in human cells and tissues, we raised antibodies in rabbits against peptides derived from the N-terminal extensions and used the affinity-purified antibodies for western blot and immunofluorescence staining analyses (Figure 3). The western blot analyses revealed that both forms of ET-H1.0 were present in human breast- and lung-cancer cells but not in HeLa cells (Figure 3A). ET-H1.0 was also detected in extracts from three human-cancer tissues and three normal tissues (Figure 3A). Although ET-H1.0 was always accompanied by H1.0, its abundance did not correlate with the observed amounts of H1.0. For example, normal breast tissue and cancerous breast tissue from the same patient contained similar amounts of H1.0, whereas the abundance of the ET-H1.0 fluctuated (Figure 3A, lanes 4–9).

Immunofluorescence staining of MCF7 cells revealed a bright-speckled staining pattern of interphase nuclei (Figures 3B and 3C), and its staining pattern was distinct from that of H1.0 (Figures 3G and 3H). During mitosis, ET-H1.0 appears to be in the vicinity of condensed chromosomes (arrows in Figures 3B and 3C). Control staining of HeLa cells, in which we were not able to detect ET-H1.0 (see above), resulted in the absence of staining (Figures 3D and 3E). The preferential nuclear location of ET-H1.0 was confirmed by fractionation of MCF7 cells into nuclear and low-speed cytosolic fractions followed by western blotting (Figure 3G). Staining via a monoclonal H1.0 antibody was distinct from that via its ET form (Figures 3G and 3H).

We found that ET-H1.0 colocalizes with splicing speckles (Figures 3I–3L). The splicing speckles, subnuclear structures identified by immunofluorescence microscopy, are thought to mirror the interchromatin granule clusters detected by electron microscopy. The most widespread belief is that the

Figure 1. Identification of Linker Histone and HMG Proteins with Extra Terminal (ET) Extensions Created upon Translation of the 5′-Untranslated Regions

(A) Localization of ET-H1.0 and ET-HMGN1 in PAGE-separated nuclear protein fractions of MCF7 cells.

(B and C) Fragmentation spectra of tryptic peptides carrying the regions from both the ET extension and the "normal" protein parts of ET-H1.0 and ET-HMGN1, respectively.

(D and E) Fragmentation spectra of tryptic peptides carrying the most N-terminally identified sequences of the ET extension of ET-H1.0 and ET-HMGN1, respectively. See [25] for explanation of peptide fragmentation.

(F) Putative primary structure of the ET proteins. Bold text indicates experimentally observed ET extension; underlined text indicates sequences of indentified peptides. pS denotes phosphoserine. Negative numbering is used for amino acid residues in the ET extensions. Arrows indicate the primary structures of the long (l) and short (s) ET-H1.0 forms.

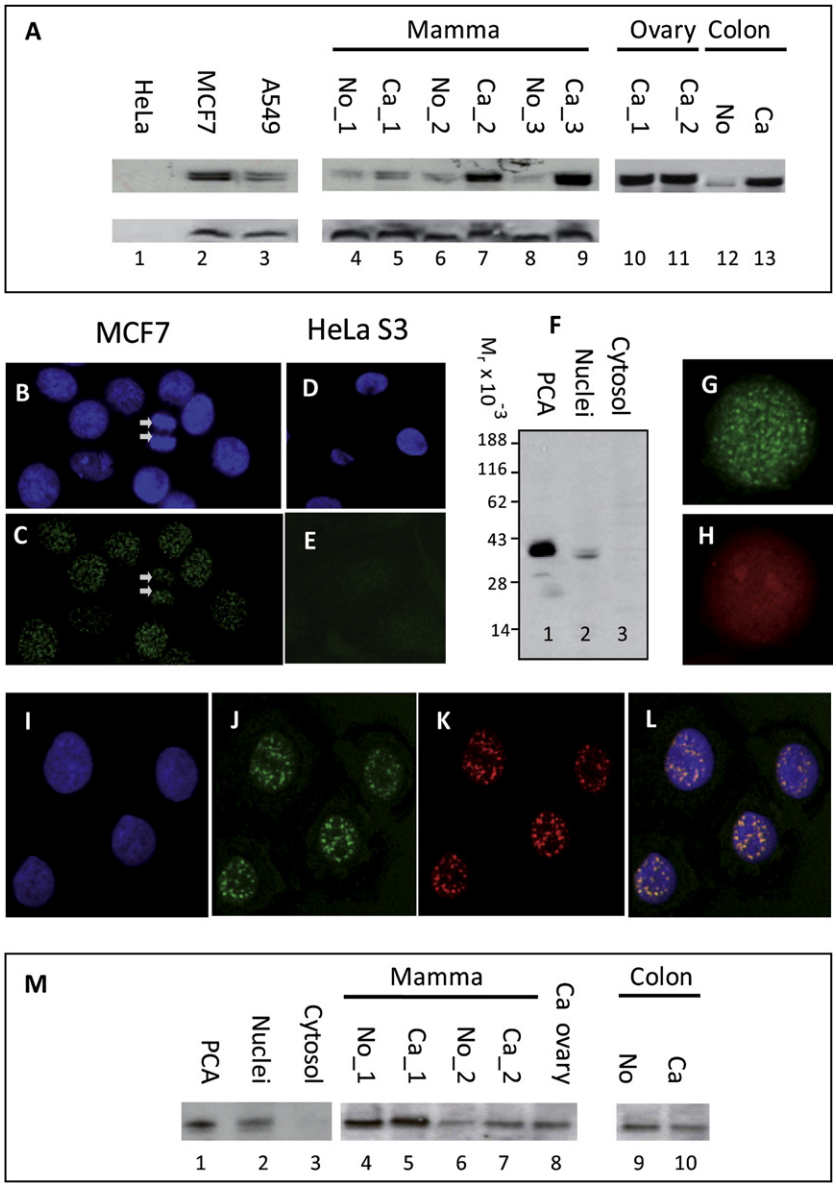


Figure 3. Occurrence of ET-H1.0 and ET-HMGN1 in Human Cells and Tissues

(A) Occurrence of ET-H1.0. Lanes 1–3, western blot analysis of perchloric acid extracts of cervical-, breast-, and lung-cancer cells, respectively. Lanes 4–9, extracts of three pairs of tissue, each matching the same patient (1, 2, or 3), of normal breast (No) and ductal invasive carcinoma G2 (Ca). Lanes 10 and 11, ovarian tumor. Lanes 11 and 13, normal colon and colon adenocarcinoma G2. The blots were probed with anti-ET-H1.0 and H1.0 antibodies.

(B–E) Nuclear localization of ET-H1.0 in MCF7 cells (B–D); absence of fluorescence in HeLa cells, which express little or no ET-H1.0 (E); fluorescent staining of nuclei with DAPI (B and D); and immunofluorescent staining with anti-ET-H1.0 antibodies (C and E).

(F) Western blot analysis of purified nuclei (lane 2) and cytosolic fraction (lane 3) of MCF7 cells. Lane 1, perchloric acid extract reference.

(G and H) Double immunofluorescent staining of the MCF7 nucleus with ET-H1.0 (green) and H1.0 (red) antibodies, respectively.

(I–L) Double immunofluorescent staining of the MCF7 nucleus with anti-ET-H1.0 (J) and SC-35 (K) antibodies, respectively; DNA staining with DAPI (I); and merged (J) and (K) (shown in [L]).

(M) Expression of ET-HMGN1. Lanes 1–3 as in (F). Lanes 4–7, extracts of two pairs, each matching the same patient (1 or 2), of normal breast (No) and ductal invasive carcinoma G2 (Ca) tissue samples. Lane 8, ovarian tumor. Lanes 9 and 10, normal colon and colon adenocarcinoma G2.

speckles serve as depots of mRNA splicing factors and supply the needed factors at the demand of the translation machinery [13]. Even so, with some reports showing that the splicing speckles could themselves operate as transcription sites [14], their function remains obscure. A number of nuclear proteins are guided to the speckles through the arginine-and-serine-rich domain [15]. It is possible that a built-in message in the ET sequence also directs ET-H1.0 toward the splicing speckles.

We also performed western blot analyses with the anti-ET-HMGN1 antibodies. This revealed that ET-HMGN1 is present in the nuclear fraction of MCF 7 cells and, like ET-H1.0, occurs in all analyzed normal and cancerous tissues (Figure 3M). We estimated the relative abundance of ET-HMGN1 compared to HMGN1 to be about 1:50 (Appendix S6).

Abundance of ET-H1.0 Generally Does Not Correlate with H1.0

Having established the existence and cellular localization of the ET proteins, we then searched for their possible regulation in biological processes. Butyrate is known to reduce the growth of many cell types, causing an arrest in the G1 phase of the cell cycle [16]. In HeLa S3 cells, the synthesis of H1.0 occurs at a very low level (mRNA and protein) but is dramatically augmented by butyrate treatment [17]. We found that in the butyrate-stimulated cells, high expression levels of H1.0 were accompanied by substantial increase of ET-H1.0. The increase was detectable on western blots (Figures 4A and 4B) and by immunofluorescence staining (Figures 4E and 4F).

Given that butyrate has been shown to induce apoptosis in human cancer cells [18], we decided to analyze the occurrence of H1.0 and its ET-H1.0 forms upon apoptosis induced by DTT [19], Vinblastine [20], and TNF-α [21]. Unlike the case of butyrate stimulation, in the cells treated with DTT, Vinblastine, and TNF-α, only ET-H1.0 was augmented. The strongest induction of ET-H1.0 was observed after treatment of the cells with DTT or TNF-α for 30 hr. In apoptotic cells, ET-H1.0 was localized in the shrunken and vacuolarized nuclei within the condensed chromatin, but the ET-H1.0-containing speckles were also observed outside the nuclei. These results suggest that the levels of ET-H1.0 are regulated independently of translation of *h1.0* gene. The nature of this regulation mechanism, as well as a potential role of ET-H1.0 in differentiation and apoptosis, remains to be elucidated. Given that the N-terminal-extension region of ET-H1.0 is highly basic (pI > 12), we speculate that it could increase the tightness of histone binding to DNA. In this way, the extension could lead to formation of more

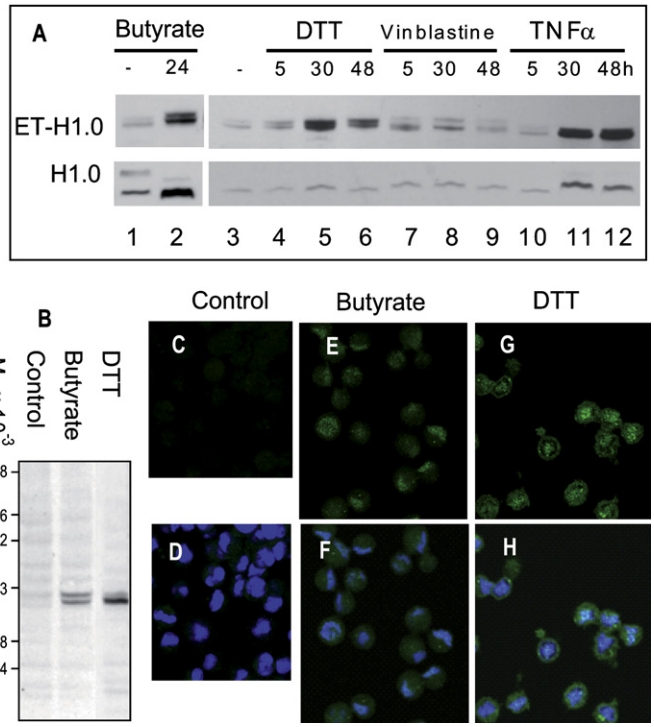


Figure 4. In HeLa S3 Cells, ET-H1.0 Can Be Stimulated by Butyrate and Other Agents, Causing Apoptosis

(A) Western blot analysis of perchloric acid extracts of HeLa S3 cells treated with butyrate, DTT, Vinblastine, and TNFα. The blots were probed with anti-ET-H1.0 and H1.0 antibodies.

(B) Western blot analysis of whole lysates of the untreated (lane 1), butyrate-treated, and DTT-treated cells (24 hr).

(C–H) Immunofluorescence analysis of cells treated with butyrate and DTT. (C), (E), and (G): anti-ET-H1.0 staining. (D), (F), and (H): merged antibody and DAPI staining.

compacted chromatin that is characteristic of both terminally differentiated and apoptotic cells.

Conclusions

Our work presents strong evidence that some 5′-noncoding sequences of human genes can be translated into proteins upon creation of the new start codon during or after transcription. The N-terminally extended proteins occur in normal human cells, and their synthesis is not related to alterations at the DNA level. For the examples described here, the amounts of ET proteins appear to be about two orders of magnitude lower than those of the standard products of the corresponding genes. However, given that H1.0 and HMGN1 occur in about 10^6–10^7 copies per cell [22–24], the ET forms are still relatively abundant compared to transcription factors and other chromatin proteins with specific regulatory functions. Presumably, the ET proteins were not identified until now because they are not predictable from the corresponding genes and occur "in the shadow" of their normal forms. Thus, in the past, the ET proteins were probably observed in many experiments but simply ignored as artifacts. The discovery of two ET proteins implies the existence of a dedicated enzymatic machinery of RNA editing, which is probably used more widely. Modern high-resolution MS data may already contain many more examples of unexpected protein forms, which could be brought to light by in-depth informatics analysis.

Experimental Procedures

The experimental procedures are described in the Supplemental Data file, available online.

Supplemental Data

Supplemental Data include Supplemental Experimental Procedures and six appendices and can be found with this article online at http://www.current-biology.com/S0960-9822(08)01294-3.

Acknowledgments

We thank Sonja Krüger for technical assistance. We are grateful to Yong Zhang and Jürgen Cox for expert bioinformatic advice. This work was supported by the Max Planck Society for the Advancement of Science, HEROIC, a 6th Framework grant of the European Commission, and the Munich Center for Integrated Protein Science (CIPSM).

Received: June 27, 2008
Revised: August 28, 2008
Accepted: September 19, 2008
Published online: November 6, 2008

References

1. Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. Nature *422*, 198–207.
2. Mann, M., Hendrickson, R.C., and Pandey, A. (2001). Analysis of proteins and proteomes by mass spectrometry. Annu. Rev. Biochem. *70*, 437–473.
3. Cravatt, B.F., Simon, G.M., and Yates, J.R., 3rd. (2007). The biological impact of mass-spectrometry-based proteomics. Nature *450*, 991–1000.
4. Kuster, B., Schirle, M., Mallick, P., and Aebersold, R. (2005). Scoring proteomes with proteotypic peptide probes. Nat. Rev. Mol. Cell Biol. *6*, 577–583.
5. Bustin, M. (1999). Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. Mol. Cell. Biol. *19*, 5237–5246.
6. Ding, H.F., Bustin, M., and Hansen, U. (1997). Alleviation of histone H1-mediated transcriptional repression and chromatin compaction by the acidic activation region in chromosomal protein HMG-14. Mol. Cell. Biol. *17*, 5843–5855.
7. West, K.L. (2004). HMGN proteins play roles in DNA repair and gene expression in mammalian cells. Biochem. Soc. Trans. *32*, 918–919.
8. Maresca, T.J., and Heald, R. (2006). The long and the short of it: linker histone H1 is required for metaphase chromosome compaction. Cell Cycle *5*, 589–591.
9. Wiśniewski, J.R., Zougman, A., Kruger, S., and Mann, M. (2007). Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue. Mol. Cell. Proteomics *6*, 72–87.
10. Boguski, M.S., Lowe, T.M., and Tolstoshev, C.M. (1993). dbEST–database for "expressed sequence tags". Nat. Genet. *4*, 332–333.
11. Gott, J.M., and Emeson, R.B. (2000). Functions and mechanisms of RNA editing. Annu. Rev. Genet. *34*, 499–531.
12. Ochsenreiter, T., Anderson, S., Wood, Z.A., and Hajduk, Z.L. (2008). Alternative RNA Editing Produces a Novel Protein Involved in Mitochondrial DNA Maintenance in Trypanosomes. Mol. Cell. Biol., in press.
13. Lamond, A.I., and Spector, D.L. (2003). Nuclear speckles: a model for nuclear organelles. Nat. Rev. Mol. Cell Biol. *4*, 605–612.
14. Moen, P.T., Jr., Johnson, C.V., Byron, M., Shopland, L.S., de la Serna, I.L., Imbalzano, A.N., and Lawrence, J.B. (2004). Repositioning of muscle-specific genes relative to the periphery of SC-35 domains during skeletal myogenesis. Mol. Biol. Cell *15*, 197–206.
15. Caceres, J.F., Misteli, T., Screaton, G.R., Spector, D.L., and Krainer, A.R. (1997). Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. J. Cell Biol. *138*, 225–238.
16. Prasad, K.N., and Sinha, P.K. (1976). Effect of sodium butyrate on mammalian cells in culture: a review. In Vitro *12*, 125–132.

17. D'Anna, J.A., Gurley, L.R., and Tobey, R.A. (1983). Extent of histone modifications and H1(0) content during cell cycle progression in the presence of butyrate. Exp. Cell Res. *147*, 407–417.
18. Chopin, V., Toillon, R.A., Jouy, N., and Le Bourhis, X. (2002). Sodium butyrate induces P53-independent, Fas-mediated apoptosis in MCF-7 human breast cancer cells. Br. J. Pharmacol. *135*, 79–86.
19. Tartier, L., McCarey, Y.L., Biaglow, J.E., Kochevar, I.E., and Held, K.D. (2000). Apoptosis induced by dithiothreitol in HL-60 cells shows early activation of caspase 3 and is independent of mitochondria. Cell Death Differ. *7*, 1002–1010.
20. Upreti, M., Lyle, C.S., Skaug, B., Du, L., and Chambers, T.C. (2006). Vinblastine-induced apoptosis is mediated by discrete alterations in subcellular location, oligomeric structure, and activation status of specific Bcl-2 family members. J. Biol. Chem. *281*, 15941–15950.
21. Wajant, H., Pfizenmaier, K., and Scheurich, P. (2003). Tumor necrosis factor signaling. Cell Death Differ. *10*, 45–65.
22. Kuehl, L., Salmond, B., and Tran, L. (1984). Concentrations of high-mobility-group proteins in the nucleus and cytoplasm of several rat tissues. J. Cell Biol. *99*, 648–654.
23. Crippa, M.P., Pash, J.M., Gerwin, B.I., Smithgall, T.E., Glazer, R.I., and Bustin, M. (1990). Expression of chromosomal proteins HMG-14 and HMG-17 in transformed human cells. Cancer Res. *50*, 2022–2026.
24. Zlatanova, J., and Doenecke, D. (1994). Histone H1 zero: a major player in cell differentiation? FASEB J. *8*, 1260–1268.
25. Steen, H., and Mann, M. (2004). The ABC's (and XYZ's) of peptide sequencing. Nat. Rev. Mol. Cell Biol. *5*, 699–711.